THE QUARTERLY JOURNAL OF EXPERIMENTAL PSYCHOLOGY, 1997, *50A* (1), 216–252

Abstraction Processes in Artificial Grammar Learning

David R. Shanks, Theresa Johnstone, and Leo Staggs
University College London, London, U.K.

Four experiments explored the extent to which abstract knowledge may underlie subjects' performance when asked to judge the grammaticality of letter strings generated from an artificial grammar. In Experiments 1 and 2 subjects studied grammatical strings instantiated with one set of letters and were then tested on grammatical and ungrammatical strings formed either from the same or a changed letter-set. Even with a change of letter-set, subjects were found to be sensitive to a variety of violations of the grammar. In Experiments 3 and 4, the critical manipulation involved the way in which the training strings were studied: an incidental learning procedure was used for some subjects, and others engaged in an explicit code-breaking task to try to learn the rules of the grammar. When strings were generated from a biconditional (Experiment 4) but not from a standard finite-state grammar (Experiment 3), grammaticality judgements for test strings were independent of their surface similarity to specific studied strings. Overall, the results suggest that transfer in this simple memory task is mediated at least to some extent by abstract knowledge.

Artificial grammar learning (AGL), like paired-associate learning in an earlier age, has become a widely used tool for the study of human learning processes. In a typical experiment, subjects are presented in the acquisition phase with strings of letters generated from a simple finite-state grammar such as that shown in Figure 1, originally created by Brooks and Vokey (1991). The grammar is entered at the left, and links are traversed until the grammar is exited at the right-hand side, and as a link is traversed a letter is picked up and added to the string. In this way, strings such as MVRVM and VXVRMVXR can be generated from the grammar shown in Figure 1. The grammar specifies certain constraints that exist in the order of string elements, much as exist in natural languages—for instance, strings can only begin with *M* or *V*. After exposure to "grammatical" training strings, subjects are informed of the existence of a set of rules

Requests for reprints should be sent to David R. Shanks, Department of Psychology, University College London, Gower St., London WC1E 6BT, U.K. Email: david.shanks@psychol.ucl.ac.uk

We thank Nick Chater, Pierre Perruchet, Martin Redington, Arthur Reber, and Mark St. John for their helpful comments. The research described here was supported in part by a project grant from the United Kingdom Biotechnology and Biological Sciences Research Council to the first author. The support of the Economic and Social Research Council is also gratefully acknowledged. The work is part of the programme of the ESRC Centre for Economic Learning and Social Evolution.

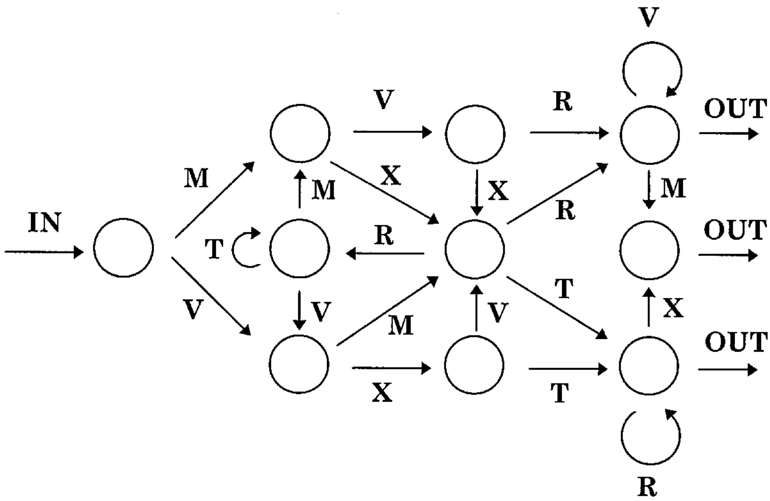

FIG. 1. The artificial grammar used by Brooks and Vokey (1991) to generate letter strings. The grammar is entered at the left-hand side and links are traversed until the grammar is exited on the right-hand side. Each link yields a new letter, which is added to the string. Strings such as *MVRVM* and *VXVRMVXR* can be generated from the grammar.

governing the structure of the training items and are then presented with a set of test items all of which are novel; some, however, are grammatical and others ungrammatical (i.e. they cannot be generated from the grammar). The subject's task is to discriminate the grammatical from the ungrammatical items.

A large number of studies have shown that subjects can perform at levels significantly above chance despite having been given minimal exposure to the study items and despite having received no instructions during the study phase regarding the later test. Typically, about 60–70% of test items are correctly classified.

Several general questions about learning have been profitably investigated in AGL studies. Amongst these are: (1) What is actually learned? (2) How is learning affected by the nature of the grammar, the task instructions, and the mental operations conducted during the study phase? (3) Can grammar learning proceed "implicitly" or unconsciously? In the present investigation we concentrate on the first two of these questions. Evidence concerning the possible unconscious status of knowledge acquired in AGL experiments has recently been reviewed by Berry and Dienes (1993) and Shanks and St. John (1994).

In recent years a number of adaptations of the basic AGL task have been explored which seem to provide evidence for two rather distinct modes of learning, which can be broadly labelled "abstractionist" and "non-abstractionist". The goal of the present studies is to investigate abstraction processes in AGL learning. By way of prelude, it is important to consider the extent to which data from AGL studies can be explained without recourse to abstraction processes. Probably the most straightforward account of what subjects learn when exposed to study items in an AGL experiment is simply the *distributional statistics* of the surface elements (i.e. the letters) from which the strings are constructed. These statistics could cover anything from simple knowledge of which letters occur at the beginning or end of strings to more sophisticated knowledge of the legal string positions of different *n-grams* (string fragments of length *n*). In the former case, the subject might simply learn that strings only ever commence with M or V, and on this basis reject novel test items beginning with other letters; in the latter case, the subject might in addition know which bigrams, trigrams, etc. are permissible and where in the string they can occur. Clearly, performance very much better than chance can be achieved simply by learning something of the statistical distribution of the surface elements from which the training strings are constructed.

What is the evidence favouring this non-abstractionist account? A number of studies (e.g. Brooks, 1978; Brooks & Vokey, 1991; Dienes, Broadbent, & Berry, 1991; Gomez & Schvaneveldt, 1994; Mathews et al., 1989; Knowlton & Squire, 1994; McAndrews & Moscovitch, 1985; Perruchet & Pacteau, 1990; Servan-Schreiber & Anderson, 1990; Vokey & Brooks, 1992) have provided support, which we will review briefly. First, and most straightforwardly, Reber and Allen (1978) asked subjects to describe their learning experience retrospectively and concurrently to justify their grammaticality judgements. Subjects reported using a variety of types of information in making their grammaticality judgements, but the violation or non-violation of expected bigrams was the most common justification, especially concerning the initial and terminal bigrams of a string. Violations of expectations about single letters—particularly the first or last letter of a string—and about trigram or longer sequences were also reported. Thus subjects plainly know a considerable amount about the statistical distribution of different n-grams in the training strings.

Perruchet and Pacteau (1990, Experiment 3) asked their subjects in the training phase to memorize strings generated from a grammar and then gave them a recognition test on letter pairs either present or absent in the training strings. Subjects performed quite well: Only 3 out of 25 old pairs were judged less familiar than any new pair, and the correlation between recognition scores and the frequency of occurrence of pairs in the training strings was 0.61. By the results of this test, then, subjects were aware of the relative frequencies of letter pairs.

In another experiment, Perruchet and Pacteau (1990, Experiment 2) took a rather different approach. They constructed test strings that contained either illegal orders of legal pairs (we call these NO items) or illegal pairs (NP items). If subjects only had information about legal pairs on which to judge the grammaticality of test strings, then strings containing illegal pairs should have been rejected, but strings comprising legal pairs in an illegal order should have been mistakenly accepted as grammatical. In accordance with this pattern, Perruchet and Pacteau found that NP items were much more likely to be rejected than NO ones. Perruchet and Pacteau then constructed a model that used pair frequency information to make grammaticality judgements. The model produced the same level of performance as subjects, except in one particular: Subjects were sensitive to the beginnings and endings of strings, but the model was not.

Finally, Perruchet and Pacteau (1990, Experiment 1) trained one group of subjects in the normal way on strings generated from the grammar and a second group on just the letter pairs comprising those strings. When subsequently required to discriminate grammatical and ungrammatical strings, the performance of the two groups was indistinguishable so long as test items containing an illegal initial letter were dropped from the analysis. Subjects trained on the letter pairs would have had no opportunity to learn which initial letters were permissible, so this procedure is not unreasonable. Perruchet and Pacteau concluded that subjects primarily knew letter pairs but also had some positional information, namely of which pairs could legally start and end strings.

The results of Perruchet and Pacteau's (1990) Experiment 1 are partly challenged by a study by Gomez and Schvaneveldt (1994). Subjects were trained on either intact strings or the bigrams or trigrams from which those strings were constructed, and ungrammatical test items contained either illegal pairs (NP) or legal pairs in illegal orders (NO). Gomez and Schvaneveldt confirmed that subjects trained on bigrams had some ability to reject NP items while being insensitive to NO violations. However, this was far exceeded by the performance of subjects trained on intact strings, who could reject both types of items. Subjects trained on trigrams fell midway between the other two groups. Although these results suggest that exposure to strings affords much more intricate learning than merely of legal letter-pairs, they do not directly challenge the view that knowledge of distributional statistics underlies artificial grammar learning: they merely suggest that the statistical knowledge encompasses more than legal letter-pairs. In the extreme, subjects could be responding to test items on the basis of similarity to whole study strings stored in memory (Brooks & Vokey, 1991).[1]

A final piece of evidence supports the view that grammaticality judgements are at least in part controlled by comparison to memorized substring (or possibly whole-item) information. In line with the general conception of so-called "instance" theories of memory (see Whittlesea, in press, for a review), it might be assumed that what is stored in memory is not simply a catalogue of statistical information about the study strings, but rather a set of "snapshots" of the study items, which preserve many aspects of their form, of the way they are mentally processed, and so on. On such a view, judgements should be susceptible to changes in the superficial characteristics of the studied strings. To test this, Whittlesea and Dorken (1993) required subjects to pronounce the training strings from one grammar and to spell the training strings from another. At test, subjects were asked either to pronounce or to spell test strings and to judge their grammatical status. Subjects were more likely to assign test strings to grammars when the encoding task matched the task for the test string than when they differed. Test strings that were equally similar to strings in both grammars were assigned to the grammar for which the encoding and test tasks matched. Such results are consistent with the idea that judgements are based on a comparison to a set of items in memory that represent the study items and their elements in a relatively unanalysed form.

[1] Recent evidence has cast doubt on the idea, proposed by Brooks and Vokey (1991), that grammaticality decisions are based on the degree of similarity between test items and whole training strings stored in memory. Knowlton and Squire (1994) failed to obtain any effect of such a similarity variable when similar and dissimilar test items were equated in terms of the frequency with which the bigrams and trigrams of which they were composed appeared in training items.

Evidence for Abstraction

Despite this support for the non-abstractionist account, three lines of evidence suggest that knowledge of n-gram statistics is insufficient to account for all the data obtained in AGL studies and, instead, support the abstractionist view that subjects in some way or other come to represent mentally the underlying grammar from which the study strings are derived. Reber (1967, 1989; Reber & Allen, 1978; Reber & Lewis, 1977) is probably the best-known proponent of this view, and the first piece of evidence in its favour comes from studies using a technique originally suggested by him in which the test items are formed using a different vocabulary from the study items.

1. Changed Letter-set Procedure. In the earliest use of the "changed letter-set" procedure, Reber (1969) trained subjects to memorize strings generated from a grammar using one set of letters (*M*, *R*, *T*, *V*, *X*). When the vocabulary was changed ($M \rightarrow W$, $R \rightarrow S$, $T \rightarrow P$, $V \rightarrow N$, $X \rightarrow Z$), subjects required fewer trials to memorize a new set of strings generated from the grammar than a new set generated from a different grammar (with either the original or the changed letters). More recently, Brooks and Vokey (1991), Gomez and Schvaneveldt (1994), Manza and Reber (in press), Mathews et al. (1989), and Whittlesea and Dorken (1993) have all found that subjects were as good or nearly as good at discriminating novel grammatical and ungrammatical strings instantiated with a new letter set as with the set used in training, and Altmann, Dienes, and Goode (1995) found that even more dramatic changes in the specific items could be tolerated. These researchers trained subjects either on strings of letters or on sequences of tones and then tested them either with novel strings in the same format as the training items or in the alternative format. Even when the surface structure of the items was changed, grammatical and ungrammatical items could be discriminated, although performance was significantly better when the items were in the same form as at study.

Taking a rather different approach, Mathews et al. (1989) trained some subjects in the standard way but asked them during the test phase to verbalize their knowledge of the grammar. They were instructed that their verbal protocols would be given to a set of yoked subjects, and that they should do their best to help these yoked subjects to discriminate grammatical from ungrammatical strings. Crucially, they were also told that their yoked partners would be seeing strings instantiated with an unknown letter set, in which case information stated purely in terms of surface features (e.g. "strings cannot begin with *Z*") would be of relatively little value. Despite this, the yoked subjects were able to perform at a level well above chance, indicating that the protocols they received must have contained information that coded abstract properties of the training strings.

All of these results suggest that exposure to strings generated from a grammar allows subjects to construct some abstract representation of the grammar. This representation is capable of being used to determine the grammatical status even of items different in surface form from the training items. For instance, subjects might learn that there are only two letters that can begin a string, that the first bigram cannot be a letter repetition, and so on. In such cases, knowledge is obviously not tied to the specific vocabulary used in training. Rather, what is learned corresponds to a feature of the abstract structure of the underlying grammar. Dienes, Altmann, and Gao (in press) have recently presented a connectionist model capable of abstracting this structure and of simulating the behavioural data. In Experiments 1 and 2 we use the changed letter-set procedure to explore in greater detail the possible abstract nature of subjects' grammatical knowledge.

2. Rule-searching Instructions. Closely allied to the evidence for transfer in changed letter-set studies is evidence that there are two distinct modes of learning that subjects can adopt in AGL studies, depending on task instructions. Reber, Kassin, Lewis, and Cantor (1980; see also Reber, 1976) found that subjects explicitly instructed to search for structural rules in the training strings were better at discriminating grammatical and ungrammatical test strings than subjects who simply memorized the training items under typical implicit learning instructions. However, this superiority extended only to situations in which the training strings were presented in a structured manner that emphasized certain aspects of their rule-governed nature. When the strings were presented in a randomized fashion, as is typically the case in AGL studies, no effect of instructions was obtained. A more detailed exploration of the effects of instructions was conducted by Mathews et al. (1989), who again found differences in the nature of acquired grammar knowledge in subjects given explicit versus implicit instructions. We discuss this study in conjunction with Experiments 3 and 4, in which we compare implicit and explicit instructions in terms of their effects on subjects' grammatical knowledge.

3. Manipulations of Similarity. A test string can be more or less similar to study items as well as being grammatical or ungrammatical. McAndrews and Moscovitch (1985), Brooks and Vokey (1991; Brooks, 1978; Vokey & Brooks, 1992), and Knowlton and Squire (1994) independently varied the grammaticality of test strings and their similarity to study items, and they found that the effects of these factors on typicality judgements were additive. Thus whether an item is grammatical per se affects the likelihood that it will be judged grammatical quite independently of its similarity to studied items (Knowlton & Squire, 1996), and on the face of it this provides another line of evidence for an abstraction process. In Experiments 3 and 4 we investigate the effects of separate manipulations of grammaticality and similarity.

The Present Studies

It is important to realize that the distinction between abstract and non-abstract knowledge is an imperfect one. We have chosen to use this terminology rather than several alternative possibilities (explicit vs. implicit, analytic vs. non-analytic, data-driven vs. conceptually driven, etc.) because it seems to us to capture best the essential variability of the mental operations subjects can deploy in AGL studies. We return in the General Discussion to considering how successful this contrast is, but for the present it is important to bear in mind that "abstract" and "non-abstract" may well represent the ends of a continuum rather than completely distinct processes. Even knowledge of distributional statistics is abstract in the sense that subjects represent *types* of n-grams: if all a subject knows is that strings can only begin with *M* or *V*, this knowledge is still abstract in the sense (1) that it applies to these letters in general (i.e. the knowledge will be perfectly well applicable to test strings written in a different font), and (2) that it may have been acquired by induction across many training strings. Instead, like Redington and Chater (1996), we regard the abstract vs. non-abstract distinction as referring to the extent to which acquired knowledge is tied to the surface form of the studied strings. When subjects can make grammaticality judgements just as well for changed-letter-set items as for items that use the same letters as the study strings, a major change in surface form is having a very small effect on performance, and the subjects' knowledge can be described as abstract. If, on the other hand, grammaticality judgements are substantially affected by a change in the operations performed on items at study and test, as in Whittlesea and Dorken's (1993) experiments, then the underlying knowledge is correspondingly non-abstract.

Another way of making the same point is in terms of the effects of similarity on grammaticality judgements. To the extent that subjects judge test items to be grammatical only when they are similar to studied strings (in terms of shared n-grams, overlapping operations performed on the study and test strings, etc.), the underlying knowledge of the grammar can be described as non-abstract. If, on the other hand, similarity to studied items plays only a minor role in determining grammaticality, then the underlying knowledge of the grammar can be assumed to be abstract. The contrast between similarity-mediated and rule-based behaviour has recently been investigated by several researchers (e.g. Allen & Brooks, 1991; Herrnstein, 1990; Regehr & Brooks, 1993; Smith & Shapiro, 1989; Smith & Sloman, 1994; Smith, Langston, & Nisbett, 1992; Ward & Scott, 1987; see Shanks, 1995, for a review).

The artificial grammar learning procedure offers a way of examing whether there is a distinction between a non-abstractionist learning system (which results in classification performance being affected by the similarity of novel test items to previous training examples) and an abstractionist system (which bases classification performance on the deep structure of the grammar). This procedure also offers a means of varying the characteristics of the rules used to construct the training and test stimuli to explore the interaction of different learning strategies with different rule systems. In the following experiments, we use each of the three manipulations described above to study the abstraction process.

EXPERIMENT 1

The objective of this experiment was to determine what kinds of information subjects abstract when they memorize representative examples that allow them to discriminate items on a classification test with a changed letter-set. To date, the only detailed evidence on this issue comes from the study by Gomez and Schvaneveldt (1994), who found that subjects trained on intact strings were able to reject test strings made ungrammatical either by the presence of an illegal bigram (NP) or by an illegal ordering of legal bigrams (NO). In either case, this is an impressive ability, because these items used an entirely different vocabulary from the training items. For instance, to reject an NP string such as *WZWZ*, the subject must be capable of mapping this on to the structure *MXMX*, which contains an illegal bigram (*XM*).

In the present study, subjects were asked to memorize representative examples of a grammar in the training phase and then to classify test items. Some of these were flawed in a specific way to allow us to identify which of five rules subjects had abstracted during training. Performance was compared between two groups who trained on the same strings but were tested on classification items in either the same (same-letters group) or a changed letter-set (changed-letters group). The performance of both of these groups was compared to that of a control group, who were trained on pseudo-randomly constructed strings and tested on the classification items in the changed letter-set.

In addition to studying a range of specifically manipulated violations, this study used a testing procedure devised by McAndrews and Moscovitch (1985), which, we hoped, would be more sensitive than that standardly used. Here, subjects were presented with pairs of strings and informed that one was grammatical and the other ungrammatical. Apart from avoiding issues of response bias, the technique of forcing subjects to compare the strings directly should maximize the utility of small amounts of fragmentary knowledge.

Method

Subjects

Thirty-six members of the general public volunteered for the experiment. They were randomly assigned to one of three groups (n = 12): the same-letters group, the changed-letters group, or the control group.

Procedure

Subjects were told that they were taking part in a memory experiment. In the first phase they were trained to remember a set of letter strings, and in the second phase they were tested on their knowledge of the grammatical structure underlying the strings they memorized during the training phase.

Training Phase

Twenty letter strings (shown in Appendix A) were presented, one at a time, on 2″ × 4″ index cards, in random order, and for approximately 5 sec each. The experimenter turned the card over, and the subject was asked to repeat the letters in the correct order. The string was re-presented if an error occurred, and the experimenter only moved on to the next card once the subject had correctly repeated the current letter string. Subjects in the same-letters and changed-letters groups memorized the Series A training strings, and subjects in the control group memorized the Series B training strings.

Test Phase

Subjects in the same-letters and changed-letters groups were told that the strings they saw during the training phase had all conformed to a complex set of rules. They were then given a test in the form of a one-page list of 25 pairs of letter strings and told that only one string in each pair conformed to the rules governing the training strings. Subjects in the control group were given the same test sheet, but they were told only that one of each pair in the test list conformed to a set of rules, with no reference being made to the training phase. All subjects were asked to indicate which letter string of each pair they believed conformed to the rules. If they were not sure, they were asked to guess, so that there was a response for every pair of strings. Subjects in the same-letters group saw strings generated using the same letter-set as their training strings, whereas subjects in the changed-letters group and the control group saw letter strings generated using a different letter-set. These last two groups were asked to try not to be confused by this.

Materials

Two series (A and B) of 20 strings were created for the training phase, and two sets of 25 string pairs were created for the test phase. Each set of strings ranged in length from five to eight characters. The 20 strings in Series A were generated from the grammar shown in Figure 1 (Brooks & Vokey, 1991; Vokey & Brooks, 1992). The strings met four criteria that ensured that they were representative of the overall grammar: all 14 of the possible bigram endings (*MR*, *MT*, *RM*, *RR*, *RV*, *RX*, *TR*, *TX*, *VM*, *VR*, *VT*, *VV*, *XR*, *XT*) were used at least once; all legal double-letter repetitions (*RR*, *TT*, *VV*) appeared at least once; the four legal initial bigrams (*MV*, *MX*, *VM*, *VX*) were used five times each; and the two legal initial letters (*M*, *V*) were paired at least once with every legal terminating letter (*M*, *R*, *T*, *V*, *X*). The Series B letter strings were all ungrammatical and were pseudo-randomly generated to ensure that all single letters and all bigrams were used evenly across all locations in the strings. The two sets of training strings are shown in Appendix A.

The 25 novel grammatical test strings were also generated from the grammar shown in Figure 1. Twenty-five ungrammatical strings were each created from a grammatical string by changing it in one of five ways: (1) an illegal bigram was created by changing the second letter to a *T*(*J*) or an *R*(*H*) (letters in brackets refer to the changed-letters group); (2) the third letter was changed to an *M*(*C*); (3) the fourth letter was changed to *X*(*N*); (4) an illegal letter repetition was placed in any position except the initial or terminal bigram; and (5) an illegal terminal bigram—*XM*(*NC*), *TM*(*JC*), or *TV*(*JL*)—was created by changing the penultimate letter. Each manipulation was applied to five grammatical strings. Random pairs consisting of a grammatical and an ungrammatical string were created so that in four pairs the strings were the same length, and in the remainder the shorter string was correct 11 times and the longer string 10 times. Within each pair, the order of the grammatical and ungrammatical strings was randomized.

Two sets of test strings were generated, one set using the same letters as in the training phase (*M*, *R*, *T*, *V*, *X*) and the other using the letters *C*, *H*, *J*, *L*, and *N*. Subjects in the same-letters group saw test strings formed from the same letters as the training strings, whereas subjects in the changed-letters and control groups saw test strings formed from a changed letter-set. The paired grammatical and ungrammatical test strings in the same letter-set are shown in Appendix B, with an indication of the violation used beside each pair, but note that subjects were presented test pairs consisting of a grammatical and ungrammatical string chosen at random from the lists. The 25 letter string pairs were presented in a random order in the classification test, but this order was the same for all subjects.

Results and Discussion

For the overall group comparison and for each violation type, planned nonorthogonal contrasts compared (1) the same-letters group with the control group, and (2) the changed-letters group with the control group. The criterion ($\alpha = 0.05/2 = 0.025$, one-tailed) was adjusted according to the Bonferroni method. Table 1 shows the mean proportions of correct responses in the classification test by type of manipulation for each of the three subject groups. Overall, subjects in the same- (mean proportion correct 0.60)

TABLE 1
Mean Proportion of Correct Classification Responses for Each Violation Type in Experiment 1

	Same-Letters Group		*Changed-Letters Group*		*Control Group*	
Violation Type	*M*	*SD*	*M*	*SD*	*M*	*SD*
1. Illegal initial bigram	0.70*	0.16	0.58	0.20	0.53	0.21
2. 3rd letter changed to M(C)	0.42	0.20	0.50	0.18	0.47	0.29
3. 4th letter changed to X(N)	0.50	0.23	0.55	0.19	0.48	0.18
4. Illegal letter repetition	0.77*	0.19	0.77*	0.17	0.50	0.20
5. Illegal terminal bigram	0.62	0.23	0.55	0.17	0.57	0.24
Overall mean proportions	0.60*	0.10	0.59*	0.06	0.51	0.10

Note: An asterisk in a same- or changed-letters group cell indicates that the result is significantly different from that of the control group.

and changed-letters (mean 0.59) groups were able to discriminate grammatical from ungrammatical strings significantly better than subjects in the control group (mean 0.51), $t(22) = 2.25$ and 2.39, respectively. The means in the same- and changed-letters groups were very similar.

The significant difference in test accuracy between the changed-letters and control groups is important because Perruchet (1994) has recently cast doubt on earlier findings with a changed letter-set. He (and also Redington & Chater, 1996) noted that previous studies (e.g. Brooks & Vokey, 1991) have contrasted the level of performance in a changed-letter-set group with the chance proportion of 0.50 rather than with the test performance of a control group trained on arbitrary strings. Such a comparison is problematic if control subjects are able to perform at a level greater than 0.50 correct, for instance as a result of learning aspects of the grammar during the test phase or as a result of biased selection of test items, which makes grammatical strings more acceptable than ungrammatical ones. In the present experiment, control subjects did not perform better than chance. Therefore, our results add to those of Altmann et al. (1995) in showing that changed letter-set subjects can indeed judge the grammaticality of test strings as a result of learning something specific about the grammar used in the training phase.

Examination of the individual violations reveals that some had a bigger impact on performance than others. There were significant differences between performance in the same-letters and control groups for Violation 1 (illegal initial bigram), $t(22) = 2.16$, and Violation 4 (invalid repetition), $t(22) = 3.37$. For the other violation types, the means of the two groups were similar and did not differ reliably, $t < 1$ in each case. The only significant difference between performance in the changed-letters and control groups was for Violation 4, $t(22) = 3.55$. For the other violation types the means of the two groups were again similar and did not differ reliably, $t < 1$ in each case.

The objective of Experiment 1 was to find out what types of rules subjects could abstract when memorizing artificial grammar strings and to see whether these rules would transfer to a classification test using different letters from the training strings. When the test was carried out with the same letters as in training, subjects were able to recognize illegal initial bigrams and illegal letter repetitions. Only the ability to recognize illegal repetitions transferred reliably to test strings created from a different letter set, and on that basis we would have to conclude that abstraction of the properties of the grammar has been restricted to just one feature. However, before drawing any firm conclusions, we report a second experiment, which used a slightly better-controlled set of test items and a still more sensitive test procedure.

EXPERIMENT 2

Detailed analysis of the test strings used in Experiment 1 reveals that in a few cases manipulations intended to violate one rule also created illegal bigrams. For example, placing an *X*(*N*) in the fourth position of a string (Violation 3) sometimes creates the only bigram the grammar cannot generate, *XM*(*NC*). To ensure that the intended structural violations were not confounded with the introduction of illegal bigrams, it was therefore necessary to design a second experiment to clarify what rules subjects learn and how well they transfer to a classification test with a changed letter set. In addition to the previous 5 violation types, a sixth manipulation was added to see whether subjects would recognize an illegal first letter. The training strings were redesigned to ensure that the six violation types always created legal bigrams, but in illegal positions, rather than illegal bigrams.

The classification test in Experiment 1 was based on subjects selecting the grammatical item in each of 25 pairs of randomly chosen strings. However, random pairing means that it is not possible to determine whether subjects have recognized the true flaw in one of the strings or whether they erroneously believe some other aspect of one of the strings to be illegal. In the present experiment, the two strings used in each of 30 classification pairs were the grammatical and flawed versions of the same string. This meant that it would be possible to identify why a string was selected as being grammatical, as the only difference between the two strings was the intentional violation, which made one string ungrammatical.

Method

Twenty-four further members of the general public were randomly assigned to the same-letters group, the changed-letters group, or the control group (n = 8). The procedure was the same as for Experiment 1.

The training strings were the same as for Experiment 1 (see Appendix A). From the grammar shown in Figure 1, 30 new grammatical test strings were generated, with lengths of five to eight letters. Ungrammatical versions of these strings were then created in one of the following six ways (see Appendix C): (1) insert an illegal first letter, (2) change the second letter to create an illegal initial bigram, (3) change the third letter to *M*(*C*), (4) change the fourth letter to *X*(*N*), (5) create an illegal letter repetition in any position except the initial or terminal bigram, (6) create an illegal terminal bigram by changing the penultimate letter. None of these manipulations created the illegal bigram *XM*(*NC*), so all test strings consisted of legal bigrams. Each of these six manipulations was applied to 5 of the 30 grammatical strings to create an ungrammatical version. The grammatical and ungrammatical versions of each string were then paired in a random order for the classification test. In addition, the order of the 30 pairs on the classification test answer sheet was randomized, although the order was the same for all subjects. Two versions of the answer sheet were created, one using the same letters as the training phase (*M*, *R*, *T*, *V*, *X*) and the other converted to the changed set of letters (*C*, *H*, *J*, *L*, *N*).

Results and Discussion

For the overall group comparison and for each violation type, results were analysed by means of two non-orthogonal planned contrasts, as in Experiment 1. Table 2 shows the mean proportions of correct responses in the classification test by type of manipulation for each of the three subject groups. Overall, subjects in the same-letter (mean proportion correct 0.77) and changed-letters (mean 0.73) groups were able to discriminate grammatical from ungrammatical strings significantly better than subjects in the control group (mean 0.44), $t(14) = 8.51$ and 6.14, respectively.

Examination of the individual violations again reveals that some had a bigger impact on performance than others, but in this experiment more of the violations yielded significant group differences. There were significant differences between performance in the same-letters and control groups for all violations except Violation 4 (4th letter changed to *X*), $t(14) > 2.20$ in each case. For Violation 4, the effect was in the expected direction, $t(14) = 1.16$. Note that the reliable effect of Violation Type 3 had not been significant in Experiment 1. Significant differences between performance in the changed-letters and control groups were obtained for Violation 2 (illegal initial bigram by changing 2nd letter), $t(14) = 3.11$, Violation 3 (3rd letter changed to C), $t(14) = 4.25$, and Violation 5 (illegal letter repetition), $t(14) = 2.97$. For each of the other violation types the means of the two groups were in the expected direction, though not statistically significant, $t(14) < 2.02$ in each case.

These results show that subjects in the same-letters group learned to recognize five of the ungrammatical flaws created by our manipulations; the only non-significant effect

TABLE 2
Mean Proportion of Correct Classification Responses for Each Violation Type in Experiment 2

	Same-Letters Group		*Changed-Letters Group*		*Control Group*	
Violation Type	*M*	*SD*	*M*	*SD*	*M*	*SD*
1. Illegal 1st letter	0.80*	0.21	0.75	0.21	0.50	0.28
2. Illegal initial bigram by changing 2nd letter	0.90*	0.11	0.75*	0.30	0.38	0.17
3. 3rd letter changed to *M*(*C*)	0.78*	0.27	0.78*	0.20	0.40	0.15
4. 4th letter changed to *X*(*N*)	0.65	0.30	0.68	0.24	0.50	0.21
5. Illegal letter repetition	0.68*	0.21	0.75*	0.18	0.40	0.28
6. Illegal terminal bigram	0.80*	0.19	0.68	0.18	0.48	0.21
Overall mean proportions	0.77*	0.08	0.73*	0.11	0.44	0.08

Note. An asterisk in a same- or changed-letters group cell indicates that the result is significantly different from that of the control group.

related to violations created by inserting an *X* in the fourth letter. Subjects in the changed-letters group were sensitive to at least three of the same rules as those in the same-letter group: illegal bigrams created by changing the second letter, an illegal third letter of C, and an illegal repetition in any position. This latter finding is in agreement with the results of Experiment 1. Overall, these results suggest that some of the rules abstracted during training in one letter set can transfer to test performance in a different letter set.

It is worth looking in some detail at the level of abstraction that is implied by the present findings. Both the same- and changed-letters groups manifested sensitivity to Violation 2, an illegal initial bigram created by changing the second letter of the test string. Although this manipulation creates an illegal string, it does not introduce any new bigrams that are illegal per se. For example, Appendix C shows that one of the ungrammatical strings created by this violation type is *VTVRM*. The two new bigrams this generates are *VT* and *TV*, but, as Appendix A shows, these are present in both lists of training strings. The same is true of the other ungrammatical strings created by this violation. The fact that subjects are sensitive to this violation implies at the very least, then, that they have abstract knowledge either of legal triplets or of the restrictions on the positioning of bigrams.

One aspect of the present results that is somewhat surprising is that subjects in the changed-letters group failed to demonstrate reliable sensitivity to the constraints on the beginnings and endings of strings, although the differences are in the right direction. This may, of course, be due to a lack of power in the analysis, but one would nonetheless expect the beginnings and endings of strings to be particularly salient (as Perruchet & Pacteau, 1990, found). Be that as it may, the results of these first two experiments confirm that—at least as indexed by performance on a changed letter-set—subjects can acquire considerable amounts of abstract knowledge of a grammar.

EXPERIMENT 3

In the next two experiments we take a different approach in order to investigate abstraction processes. As explained in the Introduction, one promising strategy is to compare the performance of subjects trained with incidental memory instructions with that of subjects explicitly asked to determine the deep structure of the grammar. Our hypothesis is that subjects trained in the latter manner will be more likely to abstract out the underlying rules of the grammar. But how can we tell whether or not subjects' knowledge is abstract? One possibility, obviously, would be to look at performance on test items with a changed letter-set, but in order to increase the scope of our findings we used a rather different procedure in these experiments. By systematically varying grammaticality and similarity to studied strings, we hoped to obtain evidence of greater knowledge of the deep structure of the grammar in subjects trained explicitly.

The procedure of independently manipulating grammaticality and similarity was first adopted by McAndrews and Moscovitch (1985). These authors arranged for each test string to differ from a given training item in one letter position or in three or more letter positions. Orthogonally, test items were either grammatical or ungrammatical. McAndrews and Moscovitch found that although more similar than dissimilar test items were called "grammatical", it was also the case that more grammatical than ungrammatical strings were classified as "grammatical" at each level of similarity—a result that was later replicated by Brooks and Vokey (1991), Knowlton and Squire (1994), and Vokey and Brooks (1992).

Unfortunately, this result is not by itself necessarily evidence of abstraction, because Perruchet (1994) and Knowlton and Squire (1994) have shown that grammatical test strings tend to contain more studied initial and final bigrams and trigrams than ungrammatical ones. Thus both the effect of similarity and the apparently independent effect of grammaticality can be reduced to substring knowledge (e.g. at the level of bigrams and trigrams). However, McAndrews and Moscovitch were able to split their subjects into two sub-groups who behaved very differently at test. Specifically, subjects in one sub-group were unaffected by the true grammatical status of test strings, whereas those in the other sub-group were unaffected by the degree of similarity between test and study strings. Regardless of the critique made by Perruchet (1994) and Knowlton and Squire (1994) of the overall group results, this more detailed pattern provides a strong hint that subjects can differ in the extent to which they abstracted the underlying structure of the grammar. In the present study, rather than doing a post hoc analysis of individual subjects' data, we manipulated the study instructions, in the hope that this would achieve a variation in the balance betwen similarity and grammaticality comparable to that obtained by McAndrews and Moscovitch. A similar procedure was adopted by Vokey and Brooks (1992), except that they attempted to increase the extent to which subjects encoded the study strings in a non-abstract form. Here, we attempt instead to encourage subjects to encode the abstract deep structure of the grammar.

We are aware of six studies that have compared the classification performance of subjects instructed to memorize instances versus those asked to learn the rules of the grammar (Dulany, Carlson, & Dewey, 1984; Perruchet & Pacteau, 1990; Reber, 1976; Reber et al., 1980; Mathews et al., 1989; Turner & Fischler, 1993). In most of these studies, instructions aimed at encouraging rule-learning were minimal (e.g. subjects were simply informed prior to a standard study phase that the strings conformed to a set of rules and that discovering these rules may be helpful). However, Mathews et al. (1989, Experiments 3 & 4; see also Mathews & Roussel, in press) developed a much more sophisticated procedure for promoting rule-discovery, and we will discuss their studies in some detail.

Mathews et al. (1989, Experiment 3) used a finite-state grammar and trained their subjects with either a *match* or an *edit* task in the study phase. The match task is similar to the standard incidental training procedure employed in Experiments 1 and 2. Subjects were shown a string of letters and asked to hold this string in memory and then to identify it in a list of five strings presented a few seconds later. The edit task, in contrast, was designed to encourage rule learning in that subjects were shown flawed examples of grammatical strings, asked to indicate which letters they thought created violations of the grammar, and then given feedback about their accuracy. Strings contained between one and four incorrect letters, and subjects adopted a hypothesis-testing strategy to determine the underlying rules used to generate grammatical strings.

In the first four blocks of the classification task of Mathews et al.'s (1989) Experiment 3, where novel items in the same letter-set were classified without feedback, subjects in both the pure match and edit groups performed at above-chance levels, and there was no difference in their performance. On the basis of this evidence, Mathews et al. suggested that finite-state grammars are difficult to learn explicitly when subjects are given instuctions to work out the underlying rules of the grammar, and that abstraction processes do not enable subjects to learn anything beyond what is automatically encoded by the non-abstractionist mechanism.

Despite this failure to dissociate the performance of match and edit groups, Mathews et al.'s procedure can be readily adapted to provide a more powerful exploration of abstraction processes. The objective of the present experiment was to manipulate independently the factors of grammaticality and similarity in the test items and look for evidence that the different training procedures can lead to a difference in the balance between the influences of similarity and grammaticality on test responding. The performance of subjects in a match group who were instructed to memorize instances was predicted to show a strong effect of similarity to training examples, because the underlying mental representations would reflect the surface form and distributional statistics of the study strings, whereas subjects in an edit group were predicted to show a much smaller similarity effect (and a stronger grammaticality effect) because the underlying representations would be the generative rules that determined grammatical string construction. The prediction, therefore, was that we would obtain an interaction between group, grammatical status, and similarity.

In Experiment 3 the training and test strings designed by Brooks and Vokey (1991) were used. Brooks and Vokey constructed four distinct types of test strings: grammatical and similar (GS) to training strings, grammatical and dissimilar (GD) to training items, ungrammatical and similar (US), and ungrammatical and dissimilar (UD). The training and test strings are shown in Appendix D. Strings were defined as similar if they were only different from a specific training string in one location and dissimilar if they were different in two or more locations. For example, it can be seen in Appendix D that the grammatical similar test item *MXRMXT* differs only by the *M* in the fourth location from the List 1 training string *MXRVXT* and that the ungrammatical similar test item *MXRRXT* differs from the same training item only by the *R* in the fourth location. All other test strings differ from *MXRVXT* by two or more locations.

Method

Subjects

Twenty-four undergraduates and postgraduates from University College London participated in the experiment and were randomly assigned to two groups (edit and match, $n = 12$). Subjects were told that they would be paid a minimum of £2 for participating, but that they could receive up to £3 for good performance. In fact, all subjects were paid £3.

Procedure

Stimulus presentation and response recording were implemented on IBM-compatible PCs with 33-cm colour monitors and standard QWERTY keyboards. The subjects in the match group were initially told that they were taking part in a short-term memory experiment, and subjects in the edit group were told they were taking part in a rule-discovery experiment. In the training phase subjects in the match group carried out 64 trials of a short-term memory matching task, and subjects in the edit group carried out 64 trials of a hypothesis-testing task. Both groups then carried out 128 trials of the same classification test.

Match Task

Subjects in the match group were told that in this phase of the experiment we would be looking at how good their short-term memory was for remembering strings of letters like *MVXTR*. On each of 64 trials, one string appeared on the screen, and the subject was asked to rehearse it mentally. The string stayed on the screen for 5 sec, and then the screen went blank for 2 sec. After this, a list of five strings was displayed, and the subject was asked to type the number (1–5) of the string that matched the one rehearsed. If the subject selected the wrong string, a beep was emitted, and the computer displayed the correct string. The 64 trials comprised four blocks of the same 16 grammatical training strings, and across blocks the 16 strings were presented in different random orders. The random orders were also different for each subject. Within the list of five strings from which subjects were asked to select the one rehearsed mentally, the correct string was placed in a random position, and again this was different across trials, blocks of trials, and subjects. The four foils were illegal versions of the correct string, with one, two, three, and four violations, respectively. The same foils were repeated in the four blocks.

Edit Task

Subjects were told that they would be shown strings of letters such as *MVXTR*, which were constructed from the five letters *M*, *R*, *T*, *V*, and *X*, and that the computer was programmed with a set of rules for putting letters into acceptable orders. Subjects were told that their task was to try to work out what these rules were. They would see one string at a time for each of 64 trials. Each string would have between one and four letters that violated the rules, in the sense that they were out of position or out of sequence. Subjects were asked to indicate whether they felt that each letter conformed to or violated the rules by putting a *Y* below letters that they believed conformed to the rules and an *N* below letters that they believed did not. It was explained that at the beginning of the experiment the subject would not know the rules, and therefore they would have to start by guessing. But on each trial they would be given feedback in the form of the correct hypothesis as a string of *Y* and *N* letters, as well as the corrected string, and they should try to learn from this feedback in order to understand the rules. They were asked to press the *X* key to go on to the next trial, once they had examined the feedback. The order of the 64 training strings was individually randomized for each subject.

Classification Task

Immediately before the classification task began, subjects in the match group were informed that the letter strings that they had been asked to memorize in the first part of the experiment were generated from a complex set of rules. They were told not to worry if they did not notice any rules, as the task they had performed made it very unlikely that they would know them. Subjects in the edit group were reminded that in the first part of the experiment they had used a hypothesis-testing strategy to try to learn the rules of the grammar. They were also told not to worry if they did not feel completely confident in their understanding of the rules, as the task was very difficult.

The 128 strings presented for classification comprised two blocks of the same 64 test strings presented in a different random order across blocks and between subjects. Each string was presented in turn, and subjects were asked to rate how well it conformed to the rules, on a scale of 1 to 6. The points on the scale indicated the following: (1) certain, (2) fairly certain, and (3) guess that the string obeys the rules; (4) guess, (5) fairly certain, and (6) certain that the string does not obey the rules.

Materials

Three separate sets of letter strings constituted the grammatical training items, ungrammatical training items, and the classification test items. The grammatical training items and all the classification test items are the same as those used by Brooks and Vokey (1991), whereas the ungrammatical training items were designed specifically for this experiment. Each of the three types of strings is described in turn.

Grammatical Training Items. The two sets of 16 grammatical training strings, for Lists 1 and 2 in Appendix D, were generated by Brooks and Vokey (1991) from the grammar shown in Figure 1. Each training item differs from every other training item on both lists by at least two letters in position. Second, the factors of item length and the use of the 24 possible node-to-node transitions of the grammar were balanced as much as possible. This was to ensure that both lists were equally representative of the grammar. The training strings were used in both the match and edit tasks. In the match task they were the items that subjects had to retain in short-term memory and then choose from a list of five items. In the edit task subjects in the edit group saw a flawed version of a training item and were eventually shown the correct grammatical training string, after they had marked where they believed there were violations. Within each group, half the subjects trained on List 1 items and the other half on List 2.

Ungrammatical Training Items. Two sets of ungrammatical training items, for Lists 1 and 2, were created. Four ungrammatical strings were constructed for each of the grammatical training items shown in Appendix D, and these four flawed items contained one, two, three, or four letter positions that violated the grammar. The violations were created using the same rules that Brooks and Vokey (1991) used to create ungrammatical test items, which they did by substituting letters: *M* with *R*, *R* with *M*, *V* with *T*, *X* with *T*, and *T* with *X*. The actual letter positions containing the violations were balanced across the seven possible locations. These flawed training strings were used in both the match and edit tasks. In the match task, subjects held a grammatical training item in short-term memory and were then asked to select it from a list of five strings, and the four related flawed items formed the four distractor strings in this task. The edit group saw a flawed string first and, after identifying the letter positions they believed were wrong, eventually saw the grammatical string. Because each grammatical string was seen four times in the edit group, each time being the corrected version of a different ungrammatical string, the two groups were equally exposed to all four ungrammatical versions of each grammatical string. The four ungrammatical foils in the match task were the same as the strings that formed the ungrammatical starting strings for subjects in the edit group. Thus both the match and edit groups saw exactly the same grammatical and ungrammatical strings in the training phase.

Classification Test Items. The 64 classification test strings are shown in Appendix D and were designed to separate the factors of grammaticality and similarity in classification performance. A test string is defined as being similar if it has only one letter position different from a training item and dissimilar if it has two or more letter positions different from every training item. Within the set of 64 test items, 16 were grammatical, and each differed from one List 1 training item by only one character (grammatical and similar), but it differed from all other training items (Lists 1 and 2) by at least two characters (grammatical and dissimilar); 16 were ungrammatical, and each differed from one List 1 training item by only one character (ungrammatical and similar), but it differed from all other training strings (Lists 1 and 2) by at least two characters (ungrammatical and dissimilar); and the same relationship holds between the same 64 test items and the List 2 training strings. Ungrammatical strings were created by substituting one letter according to the rules described above for creating flawed training strings for the edit task. The List 1 and 2 items acted as control conditions in case one set of items was easier to learn or classify than the other.

Results

Table 3 presents the mean proportions of training items on which subjects in the match and edit groups made correct responses. Within each group, the training trials were divided into four blocks of 16 trials to show changes in performance as training progressed. These data were analysed to see whether there were improvements in performance over the four blocks and whether there were differences between List 1 and List 2 training items.

Training responses given by subjects in the match group were scored as correct if the identical string as that initially presented for rehearsal was selected from the list. A two-way ANOVA for the match group, with block as a within-subjects variable and list as a between-subjects variable, indicated that there were no significant effects of block, $F < 1$, or list, $F(1, 10) = 1.26$, nor an interaction, $F(3, 30) = 1.82$. Performance was close to ceiling, and the results indicate that subjects were performing the memorization task consistently accurately across training blocks.

Subjects in the edit group were asked to indicate whether each letter in a training string was grammatical or ungrammatical by placing a *Y* or *N* beneath it. The accuracy of these responses was scored at the level of individual letters. A two-way ANOVA for the edit group indicated that there were no significant effects of block, $F(3, 30) = 2.06$, or list, $F < 1$, nor an interaction, $F(3, 30) = 1.19$. Thus, although subjects' performance did improve across blocks, the effect was not statistically reliable. Subjects did, however, perform much better than chance (0.50).

Table 4 presents the mean proportion of items classified as grammatical by subjects in the match and edit groups. Within each group, the classification responses are shown for the four test item types. The results for the match group show the same pattern across the four test item types that Brooks and Vokey (1991) observed, namely strong effects of both grammaticality and similarity in classification performance. More grammatical than

TABLE 3
Mean Proportions of Correct Responses Across Blocks in the Training Stages of Experiments 3 and 4

Experiment	*Group*	*Block 1*	*Block 2*	*Block 3*	*Block 4*	*Overall Mean*
3	Match	0.96	0.97	0.98	0.98	0.97
	Edit	0.62	0.64	0.64	0.68	0.65
4	Match	0.83	0.86	0.89	0.91	0.87
	Edit	0.66	0.75	0.78	0.81	0.75

TABLE 4
Classification Test Results in Experiments 3 and 4

Experiment	Group	*Mean Proportion "Grammatical" Responses* GS	GD	US	UD	*Mean Proportion Correct and CI*	*Mean d′ and CI*	*Mean Bias (β)*
3	Match	0.63	0.53	0.49	0.34	0.58 ± 0.02*	0.46 ± 0.12	1.08
	Edit	0.48	0.41	0.39	0.28	0.55 ± 0.03*	0.32 ± 0.18	1.22
4	Match	0.65	0.71	0.56	0.56	0.56 ± 0.06	0.34 ± 0.37	0.91
	Edit	0.70	0.70	0.24	0.20	0.74 ± 0.15*	2.14 ± 1.30	1.15
4	Edit Non-learners	0.41	0.42	0.46	0.39	0.49 ± 0.04	−0.02 ± 0.25	1.01
	Edit Learners	0.99	0.97	0.01	0.01	0.99 ± 0.02*	4.30 ± 0.37	1.00

Note: GS = grammatical and similar, GD = grammatical and dissimilar, US = ungrammatical and similar, and UD = ungrammatical and dissimilar. CI = 95% confidence interval.
* Indicates $p < .05$ against a chance level of .50.

ungrammatical items were called "grammatical", and the same held for similar and dissimilar items. The results are shown graphically in Figure 2.

Chance performance was taken as 0.50. The overall mean proportion of correct responses for the match group was 0.58, and the 95% confidence interval (0.56 to 0.60) indicates that these results are above chance. The mean proportion correct for the edit group was 0.55, with a 95% confidence interval of 0.52 to 0.58, which also suggests above-chance performance.

A three-way ANOVA comparing the proportion of items classified as grammatical (ratings < 4), with group (match or edit) as a between-subjects variable and both grammaticality and similarity as within-subjects variables, indicated that there were significant effects of grammaticality, $F(1, 22) = 51.28$, and similarity, $F(1, 22) = 34.83$. The effects of group, $F(1, 22) = 3.70$, as well as the Group × Grammaticality, $F(1, 22) = 2.64$, Group × Similarity, $F < 1$, and, crucially, Group × Grammaticality × Similarity, $F < 1$, interactions were not significant.

Because subjects in the two groups may have differed in their willingness to call strings "grammatical", the signal detection measures of relative sensitivity (d') and response bias (β) were examined. These measures were calculated for each subject and then averaged for each group (see Table 4). The sensitivity measure shows that subjects in the match group were somewhat better at discriminating grammatical from ungrammatical items than those in the edit group. Subjects in the match group showed little bias, whereas those in the edit group showed slightly more bias in favour of calling strings ungrammatical. For both groups, the level of chance responding ($d' = 0$) fell well outside the 95% confidence interval of d' scores.

Finally, subjects in the edit group were asked how well they felt they had been able to do the classification test. All of them believed that they were guessing most of the time

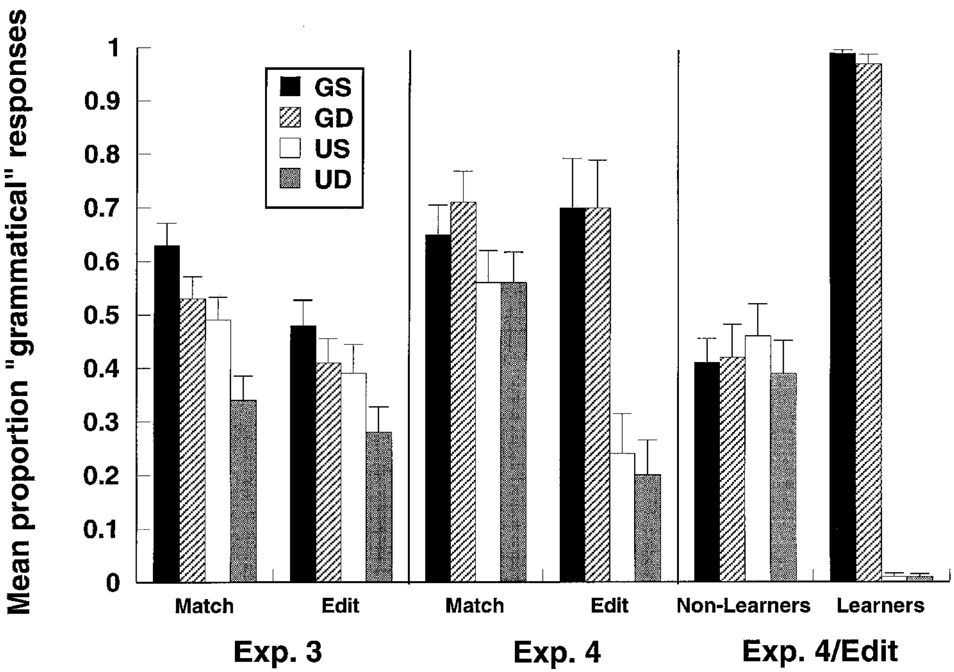

FIG. 2. Mean proportion of "grammatical" judgements for each test type and each group in Experiments 3 and 4. Data from the edit group in Experiment 4 are presented separately for learners and non-learners. GS = grammatical and similar, GD = grammatical and dissimilar, US = ungrammatical and similar, and UD = ungrammatical and dissimilar. Error bars represent standard errors.

and that they had only learned obvious rules (such as that grammatical strings always begin with *V* or *M*) and some initial bigrams and trigrams.

Discussion

On the basis of these results, it is clear that the prediction that there would be a dissociation in performance between the two groups was not supported. The match and edit groups were expected to differ in the extent to which test responding was controlled by grammaticality and similarity, with the edit group being more sensitive to grammaticality. In fact both groups showed both effects, and no difference in the balance between the factors was observable (i.e. the Group $\times$ Grammaticality $\times$ Similarity interaction was not significant).

As there was no detectable difference in the performance of the match and edit groups, no clear conclusion can be drawn about abstraction processes. One possibility is that subjects in both groups were deploying a combination of abstraction and non-abstraction processes. In this case, the obvious conclusion would be that some degree of abstraction occurs when strings are generated from a finite-state grammar even under the implicit or incidental learning conditions of the match task, and this conclusion would be consistent with the results of Experiments 1 and 2. But equally plausible (albeit at variance with the results of Experiments 1 and 2) is the possibility that performance in both groups of this experiment is entirely attributable to non-abstractionist processes. To appreciate how such an account can explain the results, we need to look in greater detail at the fact—discussed previously—that substring knowledge may underlie both the effects of grammaticality and similarity seen in this and other experiments.

While Brooks and Vokey (1991; Vokey & Brooks, 1992) set out to unconfound similarity and grammaticality, Perruchet (1994) analysed their training and test strings to show that both of these effects could be accounted for by repetition across training strings and overlap between training and test strings of initial and terminal trigrams. Perruchet's analysis is confirmed in Table 5, which shows the mean trigram frequency statistics he calculated for each of Brooks and Vokey's (1991) four types of strings.[2] Put simply, grammatical test strings contain more studied initial and terminal trigrams (mean = 3.65) than ungrammatical ones (2.77), and similar test strings (3.72) contain more studied trigrams than dissimilar ones (2.69). Merely by responding on the basis of overlap at the level of trigrams, subjects can manifest effects of both similarity and grammaticality.

An example may help to clarify the statistics presented in Table 5. For each test string, a count is made of how often its initial trigram is repeated across training strings, and this is repeated for terminal trigrams. Then the counts for initial and terminal trigrams are summed and averaged across test strings. For instance, the training strings in Appendix D show that the initial trigram *MXR* occurs at the beginning of 5 out of a total of 16 training strings, whereas the terminal trigram *RVM* occurs in 2. From this we derive a count of 5 + 2 = 7 for the test string *MXRVM*. The figures in Table 5 represent the means of these counts for each type of test item. The observed repetition across training strings is likely to provide subjects with a strong cue that strings that begin with *MXR* are grammatical. The point of these statistics is that the training and test strings used in Experiment 3 do not provide a sound basis for demonstrating abstraction. All of the results can be explained by a non-abstractionist theory. Of course, we must reiterate that such a position is at odds with the results of Experiments 1 and 2: Subjects in those studies were also trained on strings generated from a finite-state grammar, but they did show evidence of abstract knowledge.

Apart from the problem of confounding similarity and grammaticality, there is another reason why the sort of finite-state grammar used in this and previous studies may not be best suited to demonstrating abstraction processes. All of the subjects in the edit group found the training task very frustrating, and they felt that they were guessing in the classification test. They could not say what the rules of the grammar were, except for very obvious ones such as that grammatical strings always begin with *V* or *M*, and they remembered a few initial bigrams and trigrams. These difficulties may point to a reason why the explicit strategy did not yield more successful results. A characteristic of the grammar used here is that it is particularly difficult to describe a set of rules that would cover all permissible grammatical conjunctions of letters, and even if this were possible, there would be an excessively large number of these rules. The number of rules subjects

[2] The corresponding figures reported by Perruchet (1994, Table 1) contain a minor error, which we have corrected.

TABLE 5
Overlap of Initial and Terminal Fragments between Training and Test Strings in Experiments 3 and 4

		Grammatical Similar (GS)		*Grammatical Dissimilar (GD)*		*Ungrammatical Similar (US)*		*Ungrammatical Dissimilar (UD)*	
Experiment		*M*	%	*M*	%	*M*	%	*M*	%
3	Trigrams	4.16	13.00	3.13	9.78	3.28	10.25	2.25	7.03
4	Bigrams	1.36	3.78	0.97	2.69	1.44	4.00	1.19	3.31
	Trigrams	0.44	1.22	0.00	0.00	0.47	1.31	0.11	0.31

Note: Degree of overlap of test strings with training strings is shown as the mean number of times the initial and terminal trigrams (and bigrams for Experiment 4) of each test item appear in training strings, and as a percentage of the maximum possible overlap. This maximum would have occurred if every training item had the same initial and terminal trigram and these two trigrams had also started and ended all test items. The maximum for each of the four test item types would have been 16 + 16 = 32 for Experiment 3 and 18 + 18 = 36 for Experiment 4.

would need to retain explicitly would far exceed the capacity of working memory, and this suggests that a different grammar with far fewer rules might be easier for an edit group to learn explicitly and hence might yield different results.

In sum, the results of Experiment 3 indicate two factors that need to be addressed in the design of a further experiment. First, subjects in the edit group may benefit from a grammar with fewer rules to remember. Second, the design of the training and test strings must ensure that fragment repetition in training strings and overlap with test strings does not confound grammaticality and similarity manipulations.

EXPERIMENT 4

A biconditional grammar differs from a finite-state grammar in having deterministic rules that can be used to define explicitly the necessary and sufficient features of grammatical items and the relationships between these features in grammatical items. Mathews et al. (1989, Experiment 4) used a biconditional grammar based on strings of six consonants (*C*, *P*, *S*, *T*, *V*, and *X*), which were arranged in two sets of four letters separated by a dot, such as *CPST.PCVX*. There were three rules governing the relationship between letters in Positions 1 and 5, 2 and 6, 3 and 7, and 4 and 8, such that when one position contains a *C*, the other should contain a *P*, where there is an *S*, the other letter should be a *V*, and where there is a *T*, the other letter should be an *X*. It is easy to see that a subject who knows these rules can make classification decisions with 100% confidence, whereas judgements about the strings of a finite-state grammar will always tend to be probabilistic because of the family-resemblance structure of the grammatical strings.

Mathews et al. (1989, Experiment 4) used the same design as in their Experiment 3, but with a biconditional instead of a finite-state grammar. Subjects in the edit group performed above chance, but those in the match group did not. Mathews et al. (1989)

therefore obtained a dissociation in the accuracy of classification performance by their match and edit groups depending on which of two types of grammar were to be learned, and their results suggest that instance memorization is the better strategy for learning strings from a typical finite-state grammar, whereas hypothesis testing is the better strategy for learning biconditional rules.

The obvious implication is that the balance between similarity and grammaticality as controlling factors in judgements of grammaticality should be different in match and edit subjects trained on strings from a biconditional grammar. Experiment 4 used the same design as Experiment 3, except that the complex finite-state grammar was replaced with one based on a small number of deterministic rules. All training and test strings contained two sets of four consonants, separated by a central dot and created from the letters *D*, *F*, *G*, *K*, *L* and *X*. There were three biconditional rules linking letters in specific locations on each side of the dot, such that where there was a *D* in one linked position, there should be an *F* in the paired position (i.e. $D \leftrightarrow F$), *G* was paired with *L* ($G \leftrightarrow L$), and *K* was paired with *X* ($K \leftrightarrow X$), yielding strings such as *DFGK.FDLX*.

Because the number of rules has been reduced and because they are unambiguous, it was predicted that subjects in the edit group would be able to learn this grammar explicitly. However, prior research by St. John (1996; St. John & Shanks, in press) suggests that subjects in the match group are unlikely to be able to abstract the underlying rules of this grammar, as there are three intervening characters between the letters linked by each of the rules (e.g. a *D* in Letter Position 1 is linked to an *F* in Letter Position 5), and an instance memorization system may require rules spanning adjacent letter positions in order to learn the contingencies successfully.

The letter strings shown in Appendix E were specifically designed for this experiment in such a way that the problem of the confounding of grammaticality with the repetition of fragments in training and test strings was eliminated. There were two sets of training items (Lists 1 and 2) and four sets of test items divided into grammatical and similar (GS), grammatical and dissimilar (GD), ungrammatical and similar (US), and ungrammatical and dissimilar (UD). The design of the training strings ensured that no simple frequency information was provided at the level of single letters or initial and terminal fragment lengths of 4, 5, 6, 7, or 8 letters that could allow grammatical and ungrammatical test items to be discriminated. The test strings overlapped with training strings only at the level of initial and terminal bigrams and trigrams. To elaborate, the test and study strings have no overlap in terms of initial or terminal n-grams of lengths 4–8, so knowledge of which chunks of these lengths appeared in the study strings would not allow grammatical and ungrammatical test strings to be discriminated. Regarding single letters, the test lists are composed of the same letters, and each test item has maximal (100%) 1-gram overlap with the study items when overlap is computed in the way described in Table 5.

For initial and terminal bigrams and trigrams, quantitative measures of repetition and overlap are shown in Table 5. The trigram level of description can be compared with Perruchet's (1994) analysis of the strings used by Brooks and Vokey (1991) and in Experiment 3 to see that fragment frequency information has been considerably reduced. Indeed, for each test string the average number of training strings containing an initial or terminal trigram in common with it is less than 0.5. Equally important is the fact that grammatical and ungrammatical test items did not differ in terms of bigram (means 1.17 vs. 1.32, respectively) or trigram (means 0.22 vs. 0.29, respectively) overlap with the training strings. If anything, the ungrammatical test items shared slightly more fragments with the study items than did the grammatical ones.

The small amount of overlap between training and test items at the level of initial and terminal bigrams and trigrams was necessary to create the required similarity manipulation. This can be explained in relation to trigrams by comparing List 1 training items with the grammatical similar and ungrammatical similar test items to their right in Appendix E. Each test string is defined as similar because it has only two letters that differ from the appropriate training string. The actual locations of the two letters that were changed within the eight letter locations were counterbalanced across all 18 test items of each type. This means that where Positions 4 and 8 were changed, the initial trigram will remain the same as the related training item. In the same manner, where Locations 1 and 5 were changed, the terminal trigram will remain the same as in the training example. However because there is no repetition in the training strings and only this minimal overlap with the test strings, it can be seen in Table 5 that the mean and percentage fragment frequency statistics for Experiment 4 are considerably lower than for Experiment 3.

It was predicted that the results of this experiment would show a dissociation in performance between the match and edit groups. The match group should show performance close to chance because the letter locations paired by the biconditional grammar had three intervening letters, making it highly unlikely that subjects would be able to abstract the rules, and because there was insufficient fragment frequency information in the training strings to enable subjects to memorize features in training examples that predict grammaticality in the test strings. However it was predicted that subjects in the edit group would succeed in identifying the necessary and sufficient features of the rules (letter positions) and the relationships between these features (letter pairs), which would result in above-chance performance, with a grammaticality effect but no similarity effect. In sum, we anticipated finding a significant Group × Grammaticality interaction, with similarity having no influence on responding.

Method

Subjects

Twenty four new subjects from the same population, paid in the same way, were once again divided into match and edit groups (n = 12).

Procedure

The procedure was the same as for Experiment 3, except that subjects carried out 72 training trials and 144 classification trials, and a different grammar and set of strings were used. As the length of training strings in this experiment was 8 letters as opposed to an average of 6.3 in Experiment 3, subjects in the match group were given 7 sec to rehearse a string mentally while it was on the screen in the match task instead of the 5 sec given in Experiment 3.

Materials

All strings comprised two sets of four characters separated by a dot and were created from the six consonants *D*, *F*, *G*, *K*, *L* and *X* (e.g. *DFKL.FDXG*). Grammatical strings showed a particular relationship between the letters in equivalent positions on either side of the dot. There were four letter-position relationships: Position 1 with 5, 2 with 6, 3 with 7, and 4 with 8. Three rules determined which letters could be located in these paired positions: *D* must be matched with an *F*, *G* with an *L*, and *K* with an *X*. Strings were created for grammatical training items, ungrammatical training items, and four types of classification test items (see Appendix E). These were designed so that the same letter was never repeated in consecutive locations on one side of a string, and use of each of the 6 possible letters was balanced across all 8 letter locations.

Grammatical Training Strings. Two separate sets of 18 grammatical training strings (Lists 1 and 2, Appendix E) were created, and half the subjects in each group were trained on each list. These training strings were designed to ensure that each training string was different from all other strings in both lists by a minimum of four locations. Within each List, the 6 consonants appeared in each of the 8 locations three times each, and every training item had different initial and terminal letter sequences from trigrams up to 7-letter sequences. Subjects in both the match and edit groups saw these strings in the training phase: the match group was asked to memorize them, and the edit group was shown the correct grammatical string at the end of each edit trial.

Ungrammatical Training Strings. Two sets of ungrammatical training items were created, one for List 1 training items and one for List 2 (see Appendix E). For each grammatical training string, four ungrammatical versions were created containing 2, 4, 6, or 8 letters that violated the three rules that specify the relationships between letters in key positions. Subjects in both the match and edit groups saw these illegal strings in the training phase. Subjects in the match group were asked to remember a grammatical training item and then to select it from a list of 5 presented 7 sec later. The 4 related ungrammatical items were the distractors in this match task. Subjects in the edit group were presented with ungrammatical strings for hypothesis testing in the edit task.

Classification Test Strings. Seventy-two classification test strings were created, and these are also shown in Appendix E. Within these 72 items, one set of 18 strings is grammatical, and each item differs from one List 1 training item (and hence is similar to it) by only two characters in specific locations, whereas it differs from all other items in its own list and in the alternate list by a minimum of three locations. The one training item that is similar is on the same line in the second column of the table in Appendix E. In the same manner, there are 18 List 2 grammatical test items, and each one is similar to one training string and differs from all others by at least three locations.

In addition, two sets of 18 ungrammatical test items were created (each related to a List 1 or List 2 item). Again each of these test items is similar to one training string, which is on the same line and differs from all other strings by a minimum of three locations.

Results

The same data were collected as for Experiment 3, and the results are shown in the lower half of Table 3 and the middle of Table 4. Table 3 presents the mean proportion of training items on which subjects made a correct response. Within each group the training trials were grouped into four blocks of 18 trials to show changes in performance as the training phase progressed. A two-way ANOVA for the match group with block as a within-subjects variable and list as a between-subjects variable indicated that there was no significant effect of block, $F(3, 30) = 1.75$, or of list, $F(1, 10) = 1.54$, and there was no interaction, $F < 1$. The level of performance in the match group was consistently high, but not as accurate as in Experiment 3. This suggests that the increase in study time was not quite sufficient to compensate for the increase in string length.

A two-way ANOVA for the edit group yielded a significant effect of block, $F(3, 30) = 4.78$, $p < .01$, no effect of list, $F < 1$, and no interaction, $F < 1$. These results suggest, as predicted, that the edit group acquired new knowledge as training progressed by successfully identifying the rules of the grammar as a result of hypothesis testing and feedback. This contrasts with the results of Experiment 3, where there was no reliable evidence of increased learning in the edit group as the training phase progressed.

Table 4 presents the mean proportion of items classified as grammatical for each group in Experiment 4, together with the overall mean proportion correct. Within each group, the classification responses are shown for the four test item types, and these are presented graphically in Figure 2. The overall mean proportion of correct responses for the match group was 0.56, and the 95% confidence interval of 0.50 to 0.62 suggests that these results could have occurred by chance. The mean proportion correct for the edit group was 0.74, with a confidence interval of 0.59 to 0.89, indicating above-chance performance.

A three-way ANOVA comparing the proportion of items classified as grammatical (ratings < 4), with group (match or edit) as a between-subjects variable and both grammaticality and similarity as within-subjects variables, found significant effects of group, $F(1, 22) = 9.57$, grammaticality, $F(1, 22) = 13.43$, and, importantly, a Group × Grammaticality interaction, $F(1, 22) = 4.97$. The effects of similarity, $F < 1$, Group × Similarity, $F(1, 22) = 4.25$; Grammaticality × Similarity, $F(1, 22) = 3.12$; and Group × Grammaticality × Similarity, $F < 1$, were not significant.

Separate two-way ANOVAs comparing the proportion of items classified as grammatical were conducted on the data of the two groups, with both grammaticality and similarity as within-subjects variables. In the match group there was no effect of similarity, $F(1, 11) = 2.11$, nor any interaction of Grammaticality × Similarity, $F(1, 11) = 1.56$, but the effect of grammaticality was marginally significant, $F(1, 11) = 3.51$, $p = .088$. The edit group showed an effect of grammaticality, $F(1, 11) = 10.18$, whereas the effect of similarity, $F(1, 11) = 2.26$, and the Grammaticality × Similarity interaction, $F(1, 11) = 1.96$, were not significant.

The sensitivity measure d' (see Table 4) shows that subjects in the edit group were better at discriminating grammatical from ungrammatical items than those in the match group, as there is no overlap in the confidence intervals. Indeed, for the match group, the level of chance responding ($d' = 0$) fell inside the 95% confidence interval of the d' scores. In contrast, discrimination in the edit group was well above chance. Subjects in the match group showed a slight bias in favour of calling strings grammatical, whereas those in the edit group showed a small bias towards calling strings ungrammatical.

Inspection of individual subjects' performance within the edit group indicated that 6 subjects successfully identified the rules (learners) and 6 did not (non-learners). A second set of analyses was therefore conducted for these two subgroups, and the results are shown in the bottom rows of Table 4 and in Figure 2. The mean proportion correct for the non-learners was 0.49, with a d' score of -0.02 indicating chance performance. The learners had a mean proportion correct of 0.99 and a very narrow confidence interval of 0.97 to 1.00. Their mean d' score was 4.30.

Separate two-way ANOVAs were carried out for these two subgroups on the proportion of items classified as grammatical, with both grammaticality and similarity as within-subjects variables. For the non-learners there was no effect of grammaticality, $F < 1$, or similarity, $F(1, 5) = 1.18$, nor an interaction of Grammaticality × Similarity, $F(1, 5) = 4.19$. For the learners, in contrast, there was a significant grammaticality effect, $F(1, 5) = 4830.82$, with no effect of similarity, $F(1, 5) = 1.68$, and no interaction of Grammaticality × Similarity, $F < 1$. These results support our predictions as they show a strong dissociation in classification performance between subjects who successfully learned the rules versus those who had to rely on memorization of instances.

The 6 subjects in the edit group who learned the rules did so early in training. The 6 subjects who did not learn the rules found the training phase very frustrating: They tended to generate incorrect hypotheses and use the training examples to test these out, rather than looking at the feedback on each trial and asking what possible rules would account for the results.

Discussion

In this experiment there was a clear dissociation in classification test accuracy, with chance-level performance by subjects in the match group and the non-learners in the edit group but almost perfect performance by 6 further members of the edit group. These latter subjects showed a strong effect of grammaticality and no effect of similarity, suggesting that the mental representation underlying their performance was knowledge of the abstract principles of the grammar.

What design factors in Experiment 4 created the conditions for a dissociation in performance of the two groups? The following three factors may—individually or together—have been responsible: the number of rules required to specify the grammar fully, the number of intervening characters between letters linked by the rule system, and the level of fragment frequency information.

In Experiment 3, the finite-state grammar encompassed a large number of rules that would be extremely difficult to define and probably impossible to hold in working memory, whereas the number of rules in the grammar used in Experiment 4 could very easily be held in memory. This undoubtedly facilitated the performance of subjects in the edit group. Therefore the number of rules in a grammar may dictate whether active hypothesis-testing will be successful.

It is also important to acknowledge that although 6 members of the edit group were able to solve the training task very quickly, other members of this group failed. Those who failed appeared to be trying to impose arbitrary hypotheses on the training stimuli rather than acknowledging that they must start by guessing and then try to identify what hypotheses might explain the feedback. So, although hypothesis testing can be fast and efficient given the appropriate instructions and rule structure, it may not be suitable for all subjects. Some may need preliminary training in hypothesis testing before they can benefit from rule-learning instructions.

The almost significant grammaticality effect of the match group subjects in Experiment 4 suggests that they may have been slowly acquiring knowledge of the rules as a by-product of memorizing instances, but relative to the situation in Experiment 3, this slow accretion may have been hampered by rules that spanned three intervening letters. Whereas this factor had no adverse effect on hypothesis testing, previous research suggests that it may slow or eliminate the accretion of contingencies by the instance memorization system (St. John, 1996; St. John & Shanks, in press). Further research is therefore necessary to compare rule and instance learning using grammars with varying numbers of intervening characters between rule-linked positions in order to ascertain systematically what types of contingency the instance memorization system can acquire.

The third difference between the two experiments is that in Experiment 4 useful fragment frequency information was reduced to an absolute minimum in the test strings. In Experiment 3 the pattern of fragment frequency information confounded the grammaticality and similarity manipulations to such an extent that it was unclear whether the two groups had indeed memorized whole instances (creating a similarity effect) and deduced the rules of the grammar (creating a grammaticality effect) or whether both effects were caused by explicitly learning which initial and terminal letter fragments predicted grammaticality. In Experiment 4, neither group showed a similarity effect, which is important because it indicates that the mental representation underlying classification performance was not a collection of memorized fragments or instances.

The implications of the findings of this study indicate that instance-memorization and hypothesis-testing instructions recruit partially separate learning processes. Whether rule learning is successful depends on the number of rules necessary to define the grammar. It also appears that there are individual differences in the ability to perform the necessary hypothesis-testing operations, and not all subjects succeed in identifying the rules. At the same time, the non-abstractionist system may learn the contingencies generated by a rule but accretes them very slowly over a large number of trials.

GENERAL DISCUSSION

In recent years considerable effort has been spent trying to explore the extent to which the human learning faculty is best described in terms of abstracting deep structural regularities about a domain, or as accumulating data concerning the distributional statistics of the domain's surface elements. Plainly, some models of cognition demand the former (e.g. Chomsky, 1980), whereas others only require the latter (e.g. Jacoby & Brooks, 1984; Plunkett & Marchman, 1993). Given the successes of non-abstractionist models (see Whittlesea, in press, for a review), we conducted four experiments designed to elucidate the nature of abstraction processes in artificial grammar learning. Experiments 1 and 2 demonstrated that exposure to strings generated from a finite-state grammar can allow mental representations to be formed that are capable of supporting transfer to strings created out of different surface elements. The results of Experiment 2, in particular, suggest that the abstracted deep structure of the grammar is quite rich in that subjects were sensitive to a number of fairly subtle violations in changed letter-set strings.

In Experiments 3 and 4 we took a rather different approach. Here, the rationale was that instructions to memorize exemplars on the one hand and instructions to try to work out the deep rules of the grammar on the other should lead subjects to be differentially sensitive to the grammatical status of test strings and their similarity to studied strings when these are manipulated independently. However, with strings generated from a finite-state grammar, Experiment 3 failed to obtain any shift in the extent to which grammaticality and similarity controlled responding. But when the strings were formed according to the rules of a biconditional grammar (Experiment 4), a major shift was observed, with the grammaticality judgements of a subset of subjects from the edit group showing no influence whatsoever of similarity. In our view, this result represents clear evidence of abstract deep-structural knowledge. As many investigators (e.g., Allen & Brooks, 1991; Herrnstein, 1990; Regehr & Brooks, 1993; Smith, Langston, & Nisbett, 1992; Smith & Sloman, 1994; Ward & Scott, 1987) have recently noted, it is the controlling influence of surface similarity that is the hallmark of the distinction between abstract (analytic, explicit, rule-based) and non-abstract (holistic, implicit, similarity-based) knowledge.

Our primary concern in these experiments has been to explore the conditions under which abstraction takes place, and obviously the clearest evidence comes from the learners in the edit group of Experiment 4, who were trained under "explicit" learning conditions. However, a secondary question that has been the focus of much recent debate in the human learning literature is the extent to which abstraction occurs under "implicit" or incidental conditions. The results of Experiments 1 and 2 provide support for implicit abstraction, inasmuch as subjects learned under incidental memorization instructions but nevertheless showed evidence of transfer across a change of letter-set. But Redington and Chater (1996) have recently challenged the view that above-chance transfer to changed letter-set items is evidence for abstraction. They have, instead, proposed various models in which what is learned in the study phase is simply the distributional statistics of the study strings, with subjects then engaging in various "code-breaking" activities to determine the grammatical status of test strings.

Although Redington and Chater's (1996) perspective is controversial, it is worth asking to what extent the data from Experiments 3 and 4—to which Redington and Chater's critique does not apply—support the possibility of implicit abstraction. Our conclusions here are more modest. In Experiment 3, subjects in the match group judged more grammatical than ungrammatical test strings "grammatical", but this result can be attributed to a confounding of grammatical status with the overlap between study and test items in terms of sub-string components (Table 5). This confounding was eliminated in Experiment 4, but here subjects in the match group only showed a marginally significant grammaticality effect. Thus the results of Experiments 3 and 4 do not by themselves provide any compelling evidence for *implicit* abstraction of the rules of a grammar.

With respect to the finite-state grammar studied in Experiments 1–3, the interpretation we prefer is that abstraction was in fact manifest in both the transfer test of Experiments 1 and 2 and in the grammaticality judgements in the match condition of Experiment 3—but that the instructions to abstract in the edit condition had no detectable effect on the degree of abstraction. We acknowledge two other possibilities. One is that, for some unknown reason, the implicit learning procedure of Experiments 1 and 2 promoted abstraction, but the match procedure of Experiment 3 did not. The second possibility is that the transfer test is (as Redington and Chater propose) not a cast-iron guarantee of abstraction, and abstraction occurred in none of these three experiments. In either case, the apparent rule learning in Experiment 3 must then be explained in terms of specific memory for initial and final bigrams and trigrams. The abstraction that does occur (in the edit group of Experiment 4) is plainly explicit rather than implicit. Obviously, further research is needed on this topic.

Whatever the status of this secondary debate, it seems clear that rule-based and memorization processes can be distinguished in artificial grammar learning experiments. Of course, obtaining a dissociation in performance is one thing, explaining it is quite another. We have chosen to characterize the different representations capable of being constructed in AGL experiments in terms of the contrast between abstract and non-abstract knowledge, and plainly it is important to reflect on this distinction in the light of the results we have obtained. If one accepts that qualitatively distinct mental processes are required to account for the data, it is then necessary to ask how best these distinct processes can be characterized.

In our view, the great merit of the abstract/non-abstract dimension is that there are straightforward and widely agreed behavioural indices of abstraction, such as transfer across vocabularies, modalities, and so on. On the other hand, an important question that is left unresolved by positing a distinction between abstract and non-abstract processes is whether these should be considered as the end-points of a continuum, or whether they represent binary categories. Can one have mental representations that vary in their degree of abstractness, or does it only make sense to talk about abstract versus non-abstract representations? We had hoped that our results, especially in Experiment 3, would demonstrate a gradual shift in the amount of control exerted by grammaticality, consistent with a continuum conception of abstractness, but this was not observed. Thus at present this issue must remain unresolved. It is worth noting, however, that recent research in the field of concept learning has contrasted abstractionist and non-abstractionist processes on the assumption that these are quite independent. The dominant approach to concept learning over the last decade has been a non-abstractionist one in which a category decision concerning a test item is assumed to be determined by its similarity to memorized category instances (see Medin & Florian, 1992; Nosofsky, 1992), but recently it has been recognized that much of the data can be equally well accounted for by abstractionist, rule-learning models (e.g. Nosofsky, Palmeri, & McKinley, 1994). Although such rule-learning systems are assumed to be able to retrieve memorized items from memory to aid rule formation, the two processes are nevertheless viewed as quite different in character.

Whether the abstract/non-abstract dimension will ultimately prove to supply the best description of the underlying cognitive operations remains to be seen. In the meantime, further efforts to document abstraction processes are likely to prove highly profitable.

REFERENCES

Allen, S.W., & Brooks, L.R. (1991). Specialising the operation of an explicit rule. *Journal of Experimental Psychology: General*, *120*, 3–19.

Altmann, G.T.M., Dienes, Z., & Goode, A. (1995). Modality independence of implicitly learned grammatical knowledge. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *21*, 899–912.

Berry, D.C., & Dienes, Z. (1993). *Implicit learning: Theoretical and empirical issues.* Hove: Lawrence Erlbaum Associates Ltd.

Brooks, L. (1978). Nonanalytic concept formation and memory for instances. In E. Rosch & B.B. Lloyd (Eds.), *Cognition and categorization* (pp. 169–211). Hillsdale, NJ: Lawrence Erlbaum Associates, Inc.
Brooks, L.R., & Vokey, J.R. (1991). Abstract analogies and abstracted grammars: Comments on Reber (1989) and Mathews et al. (1989). *Journal of Experimental Psychology: General*, *120*, 316–323.
Chomsky, N. (1980). Rules and representations. *Behavioral and Brain Sciences*, *3*, 1–61.
Dienes, Z., Altmann, G.T.M., & Gao, S.-J. (in press). Mapping across domains without feedback: A neural network model of implicit learning. *Cognitive Science*.
Dienes, Z., Broadbent, D.E., & Berry, D. (1991). Implicit and explicit knowledge bases in artificial grammar learning. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *17*, 875–887.
Dulany, D.E., Carlson, R.A., & Dewey, G.I. (1984). A case of syntactical learning and judgment: How conscious and how abstract? *Journal of Experimental Psychology: General*, *113*, 541–555.
Gomez, R.L., & Schvaneveldt, R.W. (1994). What is learned from artificial grammars? Transfer tests of simple association. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *20*, 396–410.
Herrnstein, R.J. (1990). Levels of stimulus control: A functional approach. *Cognition*, *37*, 133–166.
Jacoby, L.L., & Brooks, L.R. (1984). Nonanalytic cognition: Memory, perception, and concept learning. In G.H. Bower (Ed.), *The psychology of learning and motivation*, Vol. 18 (pp. 1–47). New York: Academic Press.
Knowlton, B.J., & Squire, L.R. (1994). The information acquired during artificial grammar learning. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *20*, 79–91.
Knowlton, B.J., & Squire, L.R. (1996). Artificial grammar learning depends on implicit acquisition of both abstract and exemplar-specific information. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *22*, 169–181.
McAndrews, M.P., & Moscovitch, M. (1985). Rule-based and exemplar-based classification in artificial grammar learning. *Memory and Cognition*, *13*, 469–475.
Manza, L., & Reber, A.S. (in press). Representing artificial grammars: Transfer across stimulus forms and modalities. In D. Berry (Ed.), *How implicit is implicit learning?* Oxford: Oxford University Press.
Mathews, R.C., Buss, R.R., Stanley, W.B., Blanchard-Fields, F., Cho, J.R., & Druhan, B. (1989). Role of implicit and explicit processes in learning from examples: A synergistic effect. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *15*, 1083–1100.
Mathews, R.C., & Roussel, L.G. (in press). Abstractness of implicit knowledge: A cognitive evolutionary perspective. In D. Berry (Ed.), *How implicit is implicit learning*? Oxford: Oxford University Press.
Medin, D.L., & Florian, J.E. (1992). Abstraction and selective coding in exemplar-based models of categorization. In A.F. Healy, S.M. Kosslyn, & R.M. Shiffrin (Eds.), *From learning processes to cognitive processes: Essays in honor of William K. Estes*, Vol. 2 (pp. 207–234). Hillsdale, NJ: Lawrence Erlbaum Associates, Inc.
Nosofsky, R.M. (1992). Exemplar-based approach to relating categorization, identification, and recognition. In F.G. Ashby (Ed.), *Multidimensional models of perception and cognition* (pp. 363–393). Hillsdale, NJ: Lawrence Erlbaum Associates, Inc.
Nosofsky, R.M., Palmeri, T.J., & McKinley, S.C. (1994). Rule-plus-exception model of classification learning. *Psychological Review*, *101*, 53–79.
Perruchet, P. (1994). Defining the units of a synthetic language: Comment on Vokey and Brooks (1992). *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *20*, 223–228.
Perruchet, P., & Pacteau, C. (1990). Synthetic grammar learning: Implicit rule abstraction or explicit fragmentary knowledge? *Journal of Experimental Psychology: General*, *119*, 264–275.
Plunkett, K., & Marchman, V. (1993). From rote learning to system building: Acquiring verb morphology in children and connectionist nets. *Cognition*, *48*, 21–69.
Reber, A.S. (1967). Implicit learning of artificial grammars. *Journal of Verbal Learning and Verbal Behavior*, *6*, 855–863.
Reber, A.S. (1969). Transfer of syntactic structure in synthetic languages. *Journal of Experimental Psychology*, *81*, 115–119.
Reber, A.S. (1976). Implicit learning of synthetic languages: The role of instructional set. *Journal of Experimental Psychology: Human Learning and Memory*, *2*, 88–94.

Reber, A.S. (1989). Implicit learning and tacit knowledge. *Journal of Experimental Psychology: General*, *118*, 219–235.

Reber, A.S., & Allen, R. (1978). Analogic and abstraction strategies in synthetic grammar learning: A functionalist interpretation. *Cognition*, *6*, 189–221.

Reber, A.S., Kassin, S.M., Lewis, S., & Cantor, G. (1980). On the relationship between implicit and explicit modes in the learning of a complex rule structure. *Journal of Experimental Psychology: Human Learning and Memory*, *6*, 492–502.

Reber, A.S., & Lewis, S. (1977). Implicit learning: An analysis of the form and structure of a body of tacit knowledge. *Cognition*, *5*, 333–361.

Redington, M., & Chater, N. (1996). Transfer in artificial grammar learning: A re-evaluation. *Journal of Experimental Psychology: General*, *125*, 123–138.

Regehr, G., & Brooks, L.R. (1993). Perceptual manifestations of analytic structure: The priority of holistic individuation. *Journal of Experimental Psychology: General*, *122*, 92–114.

St. John, M.F. (1996). *Computational differences between implicit and explicit learning: Evidence from learning crypto-grammars*. Unpublished manuscript.

St. John, M.F., & Shanks, D.R. (in press). Implicit learning from an information processing standpoint. In D. Berry (Ed.), *How implicit is implicit learning*? Oxford: Oxford University Press.

Servan-Schreiber, E., & Anderson, J.R. (1990). Learning artificial grammars with competitive chunking. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *16*, 592–608.

Shanks, D.R. (1995). *The psychology of associative learning*. Cambridge: Cambridge University Press.

Shanks, D.R., & St. John, M.F. (1994). Characteristics of dissociable human learning systems. *Behavioural and Brain Sciences*, *17*, 367–447.

Smith, E.E., Langston, C., & Nisbett, R.E. (1992). The case for rules in reasoning. *Cognitive Science*, *16*, 1–40.

Smith, E.E., & Sloman, S.A. (1994). Similarity- versus rule-based categorization. *Memory and Cognition*, *22*, 377–386.

Smith, J.D., & Shapiro, J.H. (1989). The occurrence of holistic categorization. *Journal of Memory and Language*, *28*, 386–399.

Turner, C.W., & Fischler, I.S. (1993). Speeded tests of implicit knowledge. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *19*, 1165–1177.

Vokey, J.R., & Brooks, L.R. (1992). Salience of item knowledge in learning artificial grammars. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *18*, 328–344.

Ward, T.B., & Scott, J. (1987). Analytic and holistic modes of learning family resemblance concepts. *Memory and Cognition*, *15*, 42–54.

Whittlesea, B.W.A. (in press). The representation of general and particular knowledge. In K. Lamberts & D.R. Shanks (Eds.), *Knowledge representation in cognition*. London: UCL Press.

Whittlesea, B.W.A., & Dorken, M.D. (1993). Incidentally, things in general are particularly determined: An episodic-processing account of implicit learning. *Journal of Experimental Psychology: General*, *122*, 227–248.

Original manuscript received 4 January 1996
Accepted revision received 23 May 1996

APPENDIX A

Training Strings Used in Experiments 1 and 2

Series A	*Series B*
MVXRM	MVXMM
MXTRX	VXXVV
VMRVM	XRXTM
VXVTR	RTRMT
MVRVV	TTTRV
MVXTRR	MXTXXX
MXRMVR	VRRVVR
VMTRRR	XTXXRT
VXVRVV	RRVMVM
MXRVXT	TMMMMV
MVXRVVM	MRVXXTM
MXTRRRX	VTTXVVX
VMRVXTX	XXRRMRM
VXVRVMR	RMTRRMV
VMRVXVT	TVMTTMT
MVXRMVRV	MMRVVMRT
MXRTMVXR	VVTMRRXR
VMRTMVRM	XMVVXXMV
VXVRMVRV	RVVMTVVM
VMRTTVMT	TXTTMTTM

APPENDIX B

Test Strings Used in Experiment 1

Subjects in the same-letters group were tested on the strings shown here, whereas subjects in the changed-letters and control groups were tested on strings in which *M*, *R*, *T*, *V*, and *X* were replaced by *C*, *H*, *J*, *L*, and *N*, respectively. Italic elements are illegal with respect to the grammar. See main text for explanation of the violation types.

Grammatical Test Strings	*Ungrammatical Test Strings*	*Violation Type*
MVRVM	M*T*RVM	1
VXVRV	V*R*VRV	
MXTRRR	M*R*TRRR	
VMRVXVR	V*T*RVXVR	
MVXRVXTR	M*T*XRVXTR	
MXRVV	MX*M*VV	2
VXVTX	VX*M*TX	
VXVRVM	VX*M*RVM	
VXVRVMT	VX*M*RVMT	
MVXTRRRR	MV*M*TRRRR	
MXRVMT	MXR*X*MT	3
VMRVMR	VMR*X*MR	
MXRMVRV	MXR*X*VRV	
VMRTTVMR	VMR*X*TVMR	
VXVRMVXR	VXV*X*MVXR	
VMRMXR	VMR*XX*R	4
MVXRVMR	MVXR*MM*R	
VXVRVVVV	V*XX*RVVVV	
VMRTTMVR	VMRTT*MM*R	
VMRVV	V*XX*VV	
MVXRV	MVX*TM*	5
MVXTRX	MVXT*XM*	
MVXRVXT	MVXRV*TV*	
VXVTRRR	VXVTR*XM*	
MXRTTVXT	MXRTTV*TM*	

APPENDIX C

Test Strings Used in Experiment 2

Subjects in the same-letters group were tested on the strings shown below, whereas subjects in the changed-letters and control groups were tested on strings in which *M*, *R*, *T*, *V*, and *X* were replaced by *C*, *H*, *J*, *L*, and *N*, respectively. Italic elements are illegal with respect to the grammar. See main text for explanation of the violation types.

Grammatical Test Strings	*Ungrammatical Test Strings*	*Violation Type*
VMRVV	*R*MRVV	1
MXRMXR	*T*XRMXR	
VMRVXVR	*T*MRVXVR	
MXRMVXRM	*T*XRMVXRM	
MVXRMVXR	*X*VXRMVXR	
VXVRM	V*T*VRM	2
VMTRR	V*R*TRR	
MXRVMR	M*T*RVMR	
MXRVXVT	M*T*RVXVT	
VXVRMVR	V*R*VRMVR	
MVXTX	MV*M*TX	3
MVXTR	MV*M*TR	
MVXRVV	MV*M*RVV	
MVXTRRR	MV*M*TRRR	
MVXRVXTR	MV*M*RVXTR	
VMTRRX	VMT*X*RX	4
VXVTRX	VXV*X*RX	
MXRMVRV	MXR*X*VRV	
VXVRVMT	VXV*X*VMT	
VMTRRRRX	VMT*X*RRRX	
MXRVVVVM	MXRV*MM*VM	5
VMRVMT	VMR*XX*T	
VXVRVM	V*MM*RVM	
VMRTTVMR	VMRTT*XX*R	
VMRTVXVT	VM*XX*VXVT	
VXVRV	VXV*XV*	6
MXRVM	MXR*TM*	
MVXRVXT	MVXRV*TT*	
MXRTTMXT	MXRTTM*RT*	
VXVRVXTX	VXVRVX*VX*	

APPENDIX D

Training and Classification Test Items Generated From the Artificial Grammar in Figure 1 and Used in Experiment 3

The training and test items used in Experiment 3 were designed by Brooks and Vokey (1991) to meet four constraints, and these are described using the first List 1 training item MVRVM as an example. (1) The training items for each list are shown in alphabetical order so that it can be seen that each training string within each list has at least two character positions different from all other training items in the same list. The closest items to *MVRVM* have different letters in three positions, such as the second item *MVXTR*. (2) For each training item there is one test item that is both grammatical and similar to it, and this is shown on the same row in the second column. "Similar" is defined as having only one position containing a different letter. It can be seen that *MXRVM* conforms to the grammar shown in Figure 1 and that it only differs from the training string in Column 1 by the second letter. (3) For each training item there is one test item that is similar to it but ungrammatical, and this is shown on the same row in the third column. *MTRVM* does not conform to the grammar, but it is only different from *MVRVM* in the second letter position. (4) Except for the two test strings on the same row in Columns 2 and 3, all other test strings differ from a given training string by at least two letter positions. The two items closest to *MVRVM* are *MVR* in the List 1 grammatical test item list and *MVRVVVM* in the List 2 grammatical test item list. Therefore all the other 62 items in the total set of 64 test strings are dissimilar to the training string *MVRVM*, and this pattern is the same for every training item.

	Training Items	*Grammatical Test Items*	*Ungrammatical Test Items*
List 1	MVRVM	MXRVM	MTRVM
	MVXTR	MVXTX	MVXTT
	MXR	MVR	MTR
	MXRMVXR	MXRMVXT	MXRMVXX
	MXRTMVR	MXRTMXR	MXRTMTR
	MXRTVXT	MXRTMXT	MXRTRXT
	MXRVXT	MXRMXT	MXRRXT
	MXTRRR	VXTRRR	TXTRRR
	VMRMVRV	VMRMXRV	VMRMTRV
	VMRMVXR	VMRMVXT	VMRMVXX
	VMRMXTR	VMRVXTR	VMRTXTR
	VMRVVVV	VMRVVVM	VMRVVVR
	VMRVXVR	VMRVXVT	VMRVXVX
	VMTRRRR	VMTRRRX	VMTRRRT
	VXVRMXT	VXVRVXT	VXVRTXT
	VXVRVM	VXVRVV	VXVRVT
List 2	MVXRM	MVXRV	MVXRT
	MVXRMVR	MVXRMXR	MVXRMTR
	MVXRVMR	MVXRVMT	MVXRVMX
	MVXRVVV	MVXRVVM	MVXRVVR
	MXRMXRM	MXRMVRM	MXRMTRM
	MXRMXRV	MXRMVRV	MXRMTRV
	MXRTVMT	MXRTVMR	MXRTVMM
	MXRVVVM	MVRVVVM	MTRVVVM
	MXTRRRX	VXTRRRX	TXTRRRX
	MXTRRX	VXTRRX	TXTRRX
	VMRTMXT	VMRTVXT	VMRTTXT
	VMRVMT	VMRVMR	VMRVMM
	VMT	VMR	VMM
	VXVRMXR	VXVRMVR	VXVRMTR
	VXVT	VXVR	VXVM
	VXVTRRX	VXVTRRR	VXVTRRM

APPENDIX E

Training and Classification Test Items Generated from the Biconditional Grammar in Experiment 4

The training and test items were designed to meet four objectives. (1) Grammatical strings conform to the biconditional grammar: Letter Position 1 is linked to 5, 2 to 6, 3 to 7, and 4 to 8, such that when one position contains a *D*, the other contains *F*, where there is a *G*, the other is an *L*, and where there is a *K*, the other is an *X*. (2) The use of the 6 letters is balanced so that each letter appears 3 times in each of the 8 letter locations. (3) Each training string differs from all other training strings by at least 4 letter locations. (4) Each training item has a grammatical similar item in Column 2 and an ungrammatical similar item in Column 3 that each differ from the training item by only 2 letter positions. Each training item is different from all other test items by at least 3 letter locations.

	Training Items	*Grammatical Test Items*	*Ungrammatical Test Items*
List 1	DFGK.FDLX	LFGK.GDLX	LFGK.KDLX
	DGKX.FLXK	DLKX.FGXK	DFKX.FGXK
	DKFL.FXDG	DKGL.FXLG	DKGL.FXKG
	FDXG.DFKL	FDXL.DFKG	FDXK.DFKG
	FLDK.DGFX	FGDK.DLFX	FGDK.DKFX
	FXLD.DKGF	FKLD.DXGF	FGLD.DXGF
	GKDF.LXFD	XKDF.KXFD	XKDF.GXFD
	GLFX.LGDK	GLDX.LGFK	GLKX.LGFK
	GXKL.LKXG	GXKF.LKXD	GXKD.LKXD
	KLXD.XGKF	KLGD.XGLF	KLFD.XGLF
	KXGL.XKLG	FXGL.DKLG	FXGL.FKLG
	KDLF.XFGD	KDLX.XFGK	KDLG.XFGK
	LFDG.GDFL	LFXG.GDKL	LFXG.GDXL
	LGXF.GLKD	LGDF.GLFD	LGKF.GLFD
	LKGX.GXLK	LKGD.GXLF	LKGD.GXLD
	XDKG.KFXL	XFKG.KDXL	XLKG.KDXL
	XFLK.KDGX	XFLG.KDGL	XFLG.KDGF
	XGFD.KLDF	XKFD.KXDF	XLFD.KXDF
List 2	KXFG.XKDL	DXFG.FKDL	LXFG.FKDL
	XDGK.KFLX	FDGK.DFLX	FDGK.GFLX
	LDKF.GFXD	GDKF.LFXD	XDKF.LFXD
	GFKX.LDXK	GDKX.LFXK	GDKX.LGXK
	KFLD.XDGF	KGLD.XLGF	KXLD.XLGF
	DFXL.FDKG	DFKL.FDXG	DFKL.FDLG
	LGKD.GLXF	LXKD.GKXF	LFKD.GKXF
	XGLF.KLGD	XGDF.KLFD	XGDF.KLXD
	FGXD.DLKF	FGXL.DLKG	FGXK.DLKG
	DKLX.FXGK	DKLF.FXGD	DKLF.FXGL
	LKFG.GXDL	LKXG.GXKL	LKDG.GXKL
	FKDL.DXFG	FKDX.DXFK	FKDX.DXFL
	GLXK.LGKX	GLXF.LGKD	GLXD.LGKD
	FLGX.DGLK	KLGX.XGLK	KLGX.FGLK
	XLDG.KGFL	XLFG.KGDL	XLKG.KGDL
	GXDK.LKFX	GXLK.LKGX	GXLK.LKDX
	KDFL.XFDG	KGFL.XLDG	KXFL.XLDG
	DXGF.FKLD	DKGF.FXLD	DKGF.FGLD